

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

JP05341798

SYSTEM FOR GENERAL SPEAKER USE FOR WORD SPEECH RECOGNITION DEVICE

CITIZEN WATCH CO LTD

Inventor(s): HAMANO YASUKAZU

Application No. 04174976 , Filed 19920609 , Publish d 19931224

Abstract:

PURPOSE: To give a selection right of a call name to a user and to make it easy for him to use by selecting the database adaptive to the amount of speaker's voice features from a storage section and recognizing the words generated after the selection employing the database.

CONSTITUTION: When a user selects one of the call names provided in the device and pronounces it to a microphone 9, a call name discriminating circuit 10 operates and computes the amount of call name voice features and LPS cepstrum. The system searches for the amount of features which is similar to the computed amount from all of a speaker group recognition database 15 by a pattern matching. Having done this, a speaker group recognition circuit 11 is activated to assume the speaker group, which contains the feature amount with a smallest cepstrum distance, as the group to which the speaker belongs and it becomes the loop of the speech recognition database to be used against the voice generated later. Specifically a switch 13 is connected to a speaker group A5. A pattern matching voice recognition is performed using the speaker group A5 and a result 17 is outputted.

COPYRIGHT: (C)1993, JPO&Japio**Int'l Class:** G10L00300 G10L00300**MicroPatent Reference Number:** 001860239**COPYRIGHT:** (C) JPO

(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平5-341798

(43)公開日 平成5年(1993)12月24日

(51)Int.Cl.⁵

G 1 0 L 3/00

識別記号

5 3 1 J 8842-5H

5 2 1 N 8842-5H

庁内整理番号

F I

技術表示箇所

審査請求 未請求 請求項の数2(全 4 頁)

(21)出願番号 特願平4-174976

(22)出願日 平成4年(1992)6月9日

(71)出願人 000001960

シチズン時計株式会社

東京都新宿区西新宿2丁目1番1号

(72)発明者 浜野 康和

埼玉県所沢市大字下富字武野840番地 シ

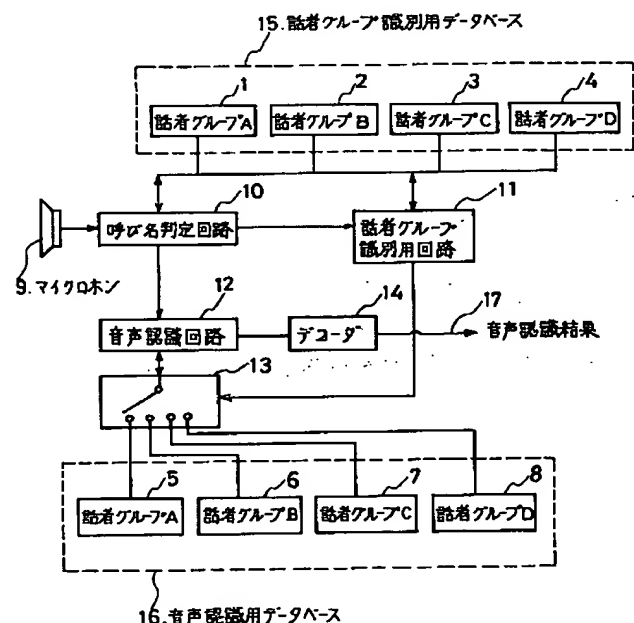
チズン時計株式会社技術研究所内

(54)【発明の名称】 単語音声認識装置の不特定話者利用方式

(57)【要約】

【目的】 100から1000単語以上の中語彙から大語彙の不特定話者単語音声認識を行う場合、ユーザに多くの単語を登録させる操作をさせることを省き、かつ、データベースをより良く構成することにより単語の認識率を高める音声認識方式を提供する。

【構成】 話者が音声認識装置につけられている呼び名の一つを発声し、その話者に適したデータベースを装置に選択させる。その後に認識させたい単語を発声し、選択されたデータベースを用いて音声認識処理が行われるようにする。



【特許請求の範囲】

【請求項 1】 不特定の話者が利用する単語音声認識装置において、音声の少なくとも一つの特徴量を基準にして構成された複数のデータベースを持ち、装置につけられた呼び名を発声することにより、話者の音声特性に適応したデータベースを記憶部分から選択し、前記データベースを用いることによりその後に発声された単語を認識することを特徴とする単語音声認識装置の不特定話者利用方式。

【請求項 2】 不特定の話者が利用する単語音声認識装置において、音声の少なくとも一つの特徴量を基準にして構成された複数のデータベースを持ち、前記複数のデータベースの中から認識すべき話者に適応したデータベースを、装置につけられた呼び名を発声することにより選択する際、前記呼び名を複数個装置に設定しておくことを特徴とする単語音声認識装置の不特定話者利用方式。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は単語音声認識装置の不特定話者利用方式に関し、特に異なる話者の音声特徴に自動的に適応して、中語彙から大語彙の単語認識を行う方式に関するものである。

【0002】

【従来の技術】 音声認識装置はマン・マシンインターフェースとして有望視されているが、その実用化は、特定話者認識装置より始まった。一人の人間だけが利用する場合には、その人の音声を入力することにより現在では数千語の単語認識が可能となってきている。ところが、複数の人間が利用する不特定話者認識装置においては、話者の音声特徴の違いのために認識が極端に困難となり、小型の装置では数十語程度しか認識できない状況にある。実用的には中語彙以上で不特定の話者を対象とした装置が望まれ、特定話者認識方式の拡張方式で対応している。

【0003】 図 3 は中語彙不特定話者認識の代表的なもののブロック図であり、登録モードと認識モードを有している。登録時においては、スイッチ 32 を登録モードにセットした後、登録したい単語をマイクロホン 33 に向かって発声し、音声登録回路 34 でその単語音声を音声特徴量、一例としては LPC ケプストラム（デジタル音声処理 p 48：古井著、東海大学出版会）に変換し、データベース 31 に蓄積される。またその単語に 1 対 1 対応する記号をキーボード等の入力装置によりデータベース 31 に登録しておく必要がある。この様に話者自信が発声した単語を登録することにより不特定話者認識においても話者あたり 100 語程度の単語を登録することができる。

【0004】 認識時においては、スイッチ 32 を認識モードにセットし、事前に登録した単語の中から認識させ

たい単語を発声させ、音声認識回路 35 でその単語の音声特徴量を算出した後、データベース 31 内の音声特徴量とパターンマッチングを行い最も類似した単語を取り出す。この際その単語に 1 対 1 対応する記号も登録されているので、その記号の方を取り出し、デコーダ 36 に送り認識結果 37 をコードとして出力する。認識時においてパターンマッチングを行う際、各話者がさまざまな単語を登録しているため、異なる単語についてその特徴量が類似する場合が生じ登録単語数が増加するにしがたが誤認識も増加する傾向にある。しかしながら、登録数が少ない場合においては話者自身の音声の特徴量をデータベースに登録しているため認識性能は高い。

【0005】

【発明が解決しようとする課題】 以上の方式によって 100 から 1000 単語以上の中語彙から大語彙の不特定話者単語音声認識を行う場合、認識性能を向上させるため、話者自身が多くの単語を登録しなければならず、ユーザに行わせる操作が多く利用しづらい点が課題となる。さらに登録単語数が増加するにつれ誤認識の割合が増え、一つのデータベースを利用する手段では限界が存在する。

【0006】

【課題を解決するための手段】 上記課題を解決するために、不特定の話者が利用する単語音声認識装置において、装置製造側で音声の少なくとも一つの特徴量を基準として多くの話者をグループ化して複数のデータベースを構成しておき、話者が装置につけられた呼び名を発声することにより、話者の音声特徴量に適応したデータベースを記憶部分から選択し、そのデータベースを用いることによりその後に発声された単語を認識する方式を特徴としている。さらに本発明においては、装置につけられている呼び名を一つに限定せず、複数個用意しておき、話者の好む呼び名を用いればよく、利用者に呼び名の選択権を与えることにより、より音声認識装置を使いやすくすることを特徴としている。

【0007】

【実施例】 以下に本発明の実施例を図面に基づいて説明する。本実施例に用いる単語音声認識装置のブロック図を図 1 に示す。機能的には、音声入力部のマイクロホン 9 と、入力された音声が発音装置につけられている呼び名かどうかを判断し呼び名である場合にその話者の音声に適したデータベースを決定する回路ブロックである呼び名判定回路 10、話者グループ識別用回路 11 と、またその際用いる話者グループ識別用データベース 15、音声認識回路 12、音声認識用データベース 16 に分けられる。ここで、話者グループ識別用データベース 15 は図の中に 1、2、3、4 で示したように、複数の話者グループ A、B、C、D に対応した話者グループ識別用のデータベースからなり、音声認識用データベース 16 は図の中に 5、6、7、8 で示したように複数の話者グルー

ブA、B、C、Dに対応した音声認識用データベースから成る。また、本システムでは単語のデータベースを持っているため従来方式に存在する登録モードは存在しない。

【0008】操作手順について説明する。話者Xがマイクロホン9に向かって装置につけられている呼び名の一つを選んで発声すると呼び名判定回路10が作動し、前記呼び名音声の特徴量、本実施例ではLPCケプストラムを算出する。この算出された特徴量に類似したものを話者グループ識別用データベース15全ての中からパターンマッチングにより探す。パターンマッチングの判定基準として本実施例ではケプストラム距離（デジタル音声処理p131：古井著、東海大学出版会）を用いたが、前記データベース中に含まれる全ての特徴量に対してケプストラム距離が基準値より大きくなった場合には呼び名を呼ばれていないとして音声認識回路12の方へ処理が移る。

【0009】基準値以内に入った場合には呼び名を呼ばれたと判断し、その後、話者グループ識別用回路11が働き前記ケプストラム距離の一番小さくなった特徴量の入っている話者グループを話者Xの属するグループとし（ここでは例として1の話者グループAとする）、その後、発声される音声に対して用いる音声認識用データベースのグループとする。具体的にはスイッチ13を5に接続する。次にマイクロホン9に入力された音声に対しても同様な処理を行うが、通常話者Xが呼び名を発声した後に目的の単語を発声するので、処理の流れは音声認識回路12へ移り、音声認識用データベース16内の5の話者グループAの部分を用いてパターンマッチング音声認識を行い音声認識結果17をデコード14により信号として出力する。

【0010】次に、話者グループ識別用データベース15と音声認識用データベース16の作成について説明する。図2に処理の流れ図を示す。まず単語音声認識に必要な単語と装置につける呼び名のリストアップを行う。ステップ21では10人以上の人間にこれらの単語を発声させ録音しておき、ステップ22で特徴量を算出する。話者グループ分けに用いる特徴量としては、本実施例ではピッチ（デジタル音声処理p8：古井著、東海大学出版会）を用いた。次にステップ23で各単語中の各母音のピッチを基にして各話者の母音[a][i]

[u][e][o]の平均値を算出しておき、それら各話者の母音の平均ピッチを本実施例では4等分することにより、ステップ24で前記の10人以上の人間が入るグループを決定する。

【0011】本実施例では4分割化を行っているが、1

6分割以下に多分割化することが望ましい。各話者グループA、B、C、Dに入る人間が決まったので、ステップ25とステップ26により話者グループ識別用データベースおよび音声認識用データベースに対して、グループ毎に各単語の認識用特徴量とその単語に1対1対応する記号を書き込んでおく。本実施例では、グループ分けに使用する特徴量にピッチを用いたが、ホルマント周波数等の他の特徴量を用いてもよい。パターンマッチングの際に用いる認識用特徴量としては本実施例では一例としてLPCケプストラムを、パターンマッチングの方式としてはDPマッチング（デジタル音声処理p162：古井著、東海大学出版会）を用いたが、他の特徴量や、他のパターンマッチング方式を用いてもよい。例えば、パターンマッチング方式にHMM法を利用し、特徴量として、ケプストラムラインやパワーを用いることもできる。

【0012】

【発明の効果】本発明の単語音声認識装置の不特定話者利用方式によれば、装置製造側においてデータベースを事前に作成しておきユーザに行わせる単語登録作業を省くことができる。またそのデータベースは話者の音声の少なくとも一つの特徴量を基準として作成した複数のものであるため、中語彙から大語彙の不特定話者認識を実現できる。さらに、装置につけられた呼び名を呼ぶことにより話者の音声の特徴を自動的に判断し、自動的に音声認識用データベースを選ぶことが可能となりユーザにとって使いやすい装置となっている。またその際、呼び名専用のデータベースが設けられているため処理スピードは高速化している。呼び名に関しては、複数用意しているため、話者に装置を利用する楽しさを与え、マン・マシンインターフェースとして自然に利用できることを可能としている。

【図面の簡単な説明】

【図1】本発明の単語音声認識装置のブロック図である。

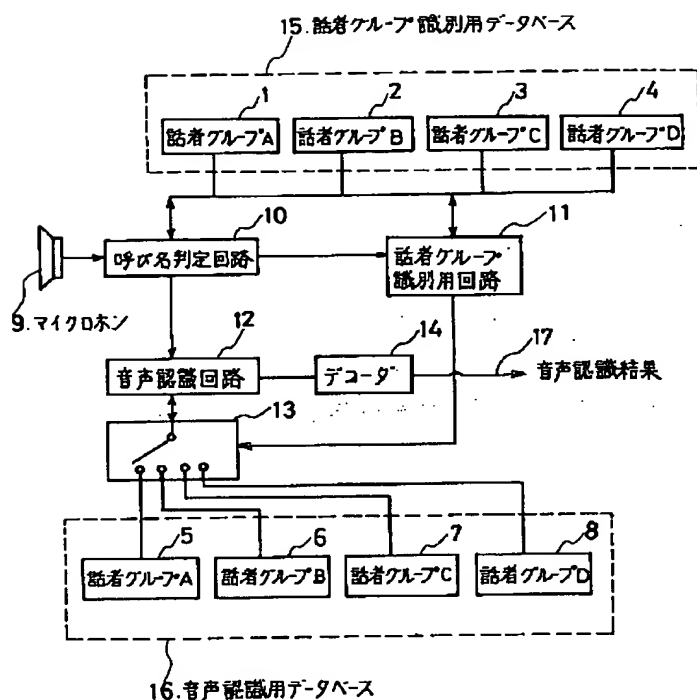
【図2】本発明の話者グループ識別用データベースと音声認識用データベースの作成の流れ図である。

【図3】従来の単語音声認識の代表的なブロック図である。

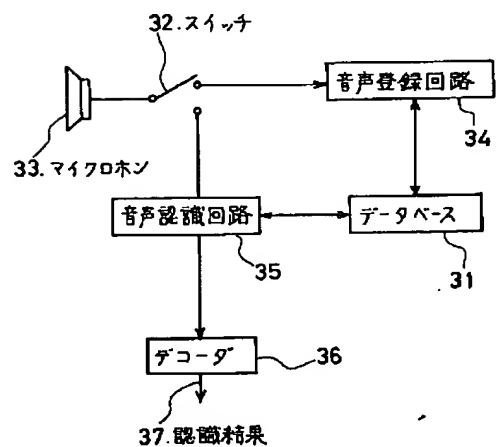
【符号の説明】

- 10 呼び名判定回路
- 11 話者グループ識別用回路
- 12 音声認識回路
- 15 話者グループ識別用データベース
- 16 音声認識用データベース

【図1】



【図3】



【図2】

